



Daniele Pasquini

Data nascita: 04/08/1989 | psqdn@gmail.com | +39 392.3286420 | Via Anassimandro 6, 00176, Roma, Italia

SKILL

PROGRAMMAZIONE

4/5 anni:

Java • Android • Python • C
R • PHP
• HTML5&CSS • JavaScript

Database:

MySQL • SQLite • PostgreSQL
• SQLServer • MongoDB

Framework:

Spark • Hadoop • Spring Web MVC •
Hibernate • CXF
• Jersey • Flask

SOFTWARE

Tableau (+ Prep) • Adobe Illustrator

SISTEMI OPERATIVI

Mac OSX • Windows (XP, Vista, 7, 8, 10)
• Linux

ISTRUZIONE

UNIVERSITÀ DI TOR VERGATA

LM IN INFORMATICA

Febbraio 2017 | Roma, Italia

Tesi in BigData e Machine Learning

Eventi su Twitter: analisi dei pattern temporali

Voto finale: 110 / 110 e lode

DATA ITALY | TOR VERGATA

FORMAZIONE AMBITO DATA ANALYTICS

Giugno 2021 | Roma, Italia

Principali corsi:

Machine Learning, Unsupervised
Learning, Supervised Learning, Text
analysis, DBMS per BigData

Progetti accademici:

classificazione non supervisionata con
clustering robusto dati sanitari,
classificazione supervisionata dati telco,
progettazione e sviluppo base di dati
ecommerce.

CORSI PRINCIPALI

MAGISTRALE

Machine Learning
Big Data e Information Retrieval
Natural Language Processing
Intelligenza Artificiale per NLP
Inferenza statistica
Algoritmi distribuiti

ESPERIENZA

WHERE TECH

DATA SCIENTIST

Lug 2021 – Presente

Analisi geopolitica su dati dal 2008 al 2020 della popolazione italiana (ISTAT), redditi famiglie per comune (Ministero dell'Economia e delle Finanze), risultati elettorali (Ministero dell'Interno). Estrazione e pulizia dei dati, unione delle banche dati, analisi descrittive mono e bivariate, analisi temporali, data viz, correlazioni e modelli predittivi. Applicativo realizzato in Python (librerie: pandas, numpy, scikit-learn, scipy, seaborn, arcgis).

MISE - TOR VERGATA

DATA SCIENTIST - RESEARCH FELLOW

Lug 2018 – Giugno 2021

Attività di ricerca in ambito Data Science:

- modelli e sistemi per l'analisi di insiemi massivi di dati di Twitter utilizzando tecniche di Machine Learning distribuito su cluster di server Hadoop con Apache Spark e MongoDB.
- ricerca e sviluppo algoritmi space efficient di conteggio approssimato tramite tecnica MinHash applicata ai grafi, calcolo del diametro effettivo e dell'average distance. Implementazione multithread in Java e Python, uso di operazioni bitwise per incremento efficienza.

DOCENZE A CONTRATTO

Ott 2019 – Presente

- Università della Tuscia (CdLM Informazione Digitale): Programmazione in Python, PySpark e SparkSQL (Apache Spark), introduzione alle principali misure statistiche e tecniche di analisi, produzione di infografiche con Tableau.
- Università di Tor Vergata (CdLM BioInformatica): Corso di programmazione e laboratorio di programmazione in C.

EVODEVO

PRESIDENZA DEL CONSIGLIO DEI MINISTRI

Mar 2020 – Mag 2020

Realizzazione di dashboard in Tableau per il Dipartimento per le Pari Opportunità su dati relativi alla pedopornografia. Pulizia dei dati con Tableau Prep.

INAIL - POC - TEXT CLASSIFIER

Nov 2019 – Dic 2019

Sviluppo di un webservice SOAP (in Python, framework Flask per il webservice, uso di scikit learn per gli algoritmi di apprendimento) in collaborazione con Oracle per la classificazione supervisionata di testi forniti in input da INAIL.

ETL PER INPS OPEN DATA

Gen 2018 – Ott 2019

Sviluppo del modulo ETL per l'estrazione e la trasformazione dei dati INPS per la creazione degli Open Data nei formati CSV, JSON, XML e OWL. Il modulo, sviluppato in Python (Pandas libreria principale), consente di gestire input/output da file in formati SAS7BDAT e CSV e database relazionali (Microsoft SQL Server, MySQL, Postgre, Oracle12c). Per la generazione degli OWL utilizzate le librerie db2triples e Apache Jena per le elaborazioni semantiche dei singoli dataset.



LINGUE

ITALIANO

Madrelingua

INGLESE

Comprensione:

Ascolto: B1, Lettura: B2

Parlato:

Interazione: B1, Produzione orale: B1

Scritto:

B2

LINK

LinkedIn:// [dpassquini](#)

ENEA ICT BIG DATA SMART STREET

Mag 2018 – Giu 2018

Installazione e configurazione dell'ecosistema Hadoop su cluster di server. Sviluppo modulo in Spark Streaming per l'acquisizione di dati da MosquittoMQTT - relativi al consumo dei lampioni installati presso la Smart Street sperimentale di Enea Casaccia - e persistenza su Data Lake HDFS.

TABLEAU CONSULTANT

Nov 2017 – Feb 2018

Consulente Tableau per Novartis. Realizzazione di dashboard per progetto pilota con dati esportati da survey online dei dipendenti e memorizzati su database Microsoft Access.

PENGEUN - EU DATATHON 2017 TOP 10

Ott 2017 – Nov 2017

Prototipo di applicazione, sviluppo algoritmo usando anche dati semantici (forniti dall'Unione Europea) per il suggerimento del miglior paese nel quale trasferirsi in base alle esigenze personali e lavorative dell'utente.

A STUDY ON SOCIAL NETWORK ANALYSIS

Sett 2017

Studio sulla propagazione dell'informazione sui social network, sul ruolo degli influencer nel processo di diffusione e sull'individuazione dei social bot all'interno delle reti.

DEMST

Mag 2018 – Lug 2018

Realizzazione di DEMST, per progetto finanziato Maites, per il tracciamento tramite device Android di veicoli e il caricamento e la persistenza dei dati su server. Sviluppo dell'app Android usando i principali componenti dell'SDK, le Google Maps API e SQLite; sviluppo di un web service RESTful in Jersey per la gestione dell'utenza, il caricamento dei dati su server e la loro persistenza su database Oracle 12c.

RECOMMENDATION SYSTEM PER INPS

Ago 2016 – Feb 2017

Ho contribuito all'ideazione e allo sviluppo di "Ti potrebbe interessare", Recommendation System all'interno di MyINPS. Applicato alle sezioni News e Servizi, il sistema profila gli utenti ed è in grado di fornire ad essi contenuti personalizzati, in base ad un ranking stilato a partire dalle interazioni con il portale. Sviluppo delle stored procedure in SQL Server per l'estrazione dei dati e calcolo del ranking dei contenuti per cluster di utenti.

OPEN DATA DASHBOARD PER INPS

Set 2016 - Dic 2016

Realizzazione del Cruscotto Gestione Open Data di INPS, per gestire il processo di creazione degli Open Data, per il linking dei metadati ai dataset e per la creazione degli scheduling per il processo di trasformazione dei dataset. Webapp realizzata in Spring MVC con Hibernate lato backend (database MySQL), HTML5, CSS con Bootstrap, Ajax e JavaScript lato frontend.

ENEA SIAPE

Nov 2015 – Mar 2016

Realizzazione di SIAPE, catalogo nazionale degli Attestati di Prestazione Energetica (APE) per ENEA. Sviluppo del servizio SOAP con Apache CXF per l'invio di APE sia in modalità individuale che con invii multipli.

- Sviluppo del database PostgreSQL, usando JAXP per la conversione degli APE in XML in oggetti Java e Hibernate per la persistenza e l'interrogazione dei dati. Ho lavorato all'autenticazione degli utenti tramite OpenLDAP.



FREELANCE | FULL STACK DEVELOPER

Gen 2020 – Gen 2021 | Roma, Italia

Sviluppo sito web (HTML5, CSS, Javascript) per una competizione dell'ente di promozione sportiva CSEN Nazionale e applicazione web (framework Flask) per l'iscrizione degli atleti alla competizione.

Sviluppo sito web (HTML5, SASS CSS, Javascript, Wordpress) per struttura ricettiva a Nettuno.

FONDAZIONE UGO BORDONI - BIGDATA LAB | DATA SCIENCE INTERNSHIP

Gen 2016 – Ago 2016 | Roma, Italia

Supervisor: Dr. Giambattista Amati, Prof. Giorgio Gambosi

Analisi, individuazione e classificazione di eventi unici su Twitter. Gestione dati in ambiente distribuito con Hadoop (HDFS e Yarn) v. 2.7 e Spark Python/Scala (con Spark SQL e Dataframe) v 1.5.2 su cluster di 8 server.

Analisi statistiche con R e Python (ggplot2 per i grafici), uso di algoritmi di Machine Learning non supervisionato per individuazione topic, estrazione di concetti con Wordnet distribuito e su grafo delle risposte di Twitter con Spark GraphX.

PUBBLICAZIONI

- Giambattista Amati, Simone Angelini, Giorgio Gambosi, Daniele Pasquini, Gianluca Rossi, Paola Vocca: Twitter: temporal events analysis: Extended Abstract. GOODTECHS 2018: 298-303
- Giambattista Amati, Simone Angelini, Anna Caterina Carli, Giorgio Gambosi, Daniele Pasquini, Giuseppe Pierri, Gianluca Rossi, Paola Vocca: Analisi temporale degli eventi su Twitter. La Comunicazione, Volume LXI: 29-42
- Antonio Cruciani, Daniele Pasquini, Paola Vocca, Giambattista Amati: About Graph Index Compression Techniques. IIR 2019, September 16–18, 2019, Padova, Italy

CERTIFICAZIONI

- Neural Networks and Deep Learning a cura DeepLearning.ai, su Coursera. Certificazione ottenuta il 12 Novembre 2017
- Improving Deep Neural Networks: Hyperparameter tuning, Regularization and Optimization a cura DeepLearning.ai, su Coursera. Certificazione ottenuta il 24 Dicembre 2017 ^{Testo}
- Introduction to Big Data a cura di University of California, San Diego su Coursera. Certificazione ottenuta l'8 Gennaio 2016
- Hadoop Platform and Application Framework a cura di University of California, San Diego su Coursera. Certificazione ottenuta il 30 Gennaio 2016
- Machine Learning With Big Data a cura di University of California, San Diego su Coursera. Certificazione ottenuta il 13 Febbraio 2016
- Graph Analytics for Big Data a cura di University of California, San Diego su Coursera. Certificazione ottenuta il 6 Marzo 2016
- MongoDB for Developers a cura di MongoDB, Inc. Attestato ottenuto a Febbraio 2016

